International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 20th Jan 2021. Link :http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-01

DOI: 10.48047/IJIEMR/V10/I01/38

Title **ANALYSIS OF AI AND GENETIC ALGORITHM UNDER CYBER SECURITY**

Volume 10, Issue 01, Pages: 188-198.

Paper Authors

CH N Santhosh Kumar, V. Sitaramu, K. Sudheer Reddy

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ANALYSIS OF AI AND GENETIC ALGORITHM UNDER CYBER SECURITY

*CH N Santhosh Kumar, Professor , Anurag Engineering College, Kodada, India.
santhosh.ph10@gmail.com
** V. Sitaramu, Professor, Swarna Bharathi Institute of Science and Technology, Khammam, India
vsitaramu.1234@gmail.com
*** K. Sudheer Reddy, Professor, Anurag Group of Institutions , Hyderabad, India.
sudheercse@gmail.com

ABSTRACT

The number of attacks is on the increase, with cyber security being a key risk for any company. Growing cybersecurity attacks endanger our lives. Artificial (AI) and computer education (ML) are capable of detecting threats and offering cyber analysts suggestions. Advancing cybersecurity-applied AI / ML includes collaborations between industries, academia and public authorities worldwide. The emphasis of this paper is on improving machine learning in which various types of computer processes in a multi-tasking environment can be mapped. To learn and characterise the cyber-consciousness of a computer process against various concurrent threads, a SHOWAN software mapping and modelling paradigm is developed. The method examined started to outperform and handled multiple tasks badly, but in connexion with anomaly detection it eventually learnt to acquire and handle tasks.

1. INTRODUCTION

The internet is becoming more and more an every second part of our lives. Every day, a new development happens, making the current system redundant. It's not always easy to adapt to this transition. The Internet's threats are various and have an important impact on the safety of users. Every process is automated with the advancement of artificial intelligence and machine learning. Artificial knowledge and machine learning are useful for internet users but also for hackers using AI to orchestrate many cyber attacks.

Figure 1: Cybersecurity hacking diagram.

What is Cybersecurity?

Cybersecurity is a defence against identity theft, software or hardware harm, and other intellectual property, of computers or related devices. Cybersecurity is critical and significant, as data is sponsored by all sectors of society, including governments , businesses, the military, various financial institutions, etc. They store vast quantities of data on computers and other internet-connected devices. These data are often combined from classified information which can not be accessed and used by the public. Those data sources are quite widely used to share information across the network, exposing the data to many cyber risks. Any third party can misuse these data easily. The most important concern of all internet users is safety, as we all continue to feed a certain amount of data to our smart phone's and personal computers.

You can enrol for free Live Demo Machine Learning online training for in-depth machine learning

Cyber-attacks may lead to the following:

- Stealing of identity, extortion of information that could lead to blackmail
- Induction of malware into networks that affect many networks by injecting viruses
- Spoofing, spamming and phishing
- Denial of different services that can lead to further attacks
- Theft of password
- Critical knowledge sabotage
- Vandalism on different websites
- Web browser privacy misuse
- Money schemes and account hacks
- Ransomware Ransomware
- Intellectual property stealing
- Illegal device and laptop access

Cybersecurity tries to avoid identity theft, multiple privacy infringements and attacks on malware and ransomware. The best way to avoid fraud online is to help mitigate risk. Either alone or with the aid of a third person who specialises in the field can handle cybersecurity. Cyber security protects different organisations from malware , phishing, ransomware and social engineering with effective steps to deter cyber-attacks.

What is Artificial Intelligence and Machine Learning?

Artificial intelligence and machine learning are data based approaches to decision-making without an explicit programming. Processes are automated by means of artificial intelligence, so that the company operations are free from human interference and discrimination.

The way businesses take decisions is influenced by artificial intelligence. This encourages computers to perform their own jobs, which had been achieved historically by hiring workers to run different computers. Data and the algorithm are provided as the input by application of AI to teach the computer to perform a particular task at maximum accuracy. Processes are streamlined with the aid of AI and tasks are simple and error-free. Data are often derived and different patterns based on past trends based on artificial intelligence and machine learning. These patterns help to determine the future and the present.

Effect of AI and ML on Cyber Security

New methodologies for automating and risk-free cyber security space are being built through development in the field of artificient intelligence and the rise in the number of apps for machine learning. The cybersecurity workers can easily coordinate and handle log data with the implementation of these elements. There are a variety of data points which can use artificial intelligence because cybersecurity requires all data clustering, ranking, sorting, filtering and handling.

AI can not develop and operate on its own, although it is a very strong idea. It requires unique chunks of data, which should be the basis for decisions. Machine Learning analyses historical data and then provides the perfect solutions for the present and for the future. For the purposes of hybrid machine learning , artificial intelligence and cyber security, previous data must be made available. Therefore.

In order to organise past data efficiently, the algorithms must be given. The device must then include instructions for different elements and patterns on which threats and other malware can be screened. The algorithms must be built to distinguish easily between a normal situation and a situation where there are compromises on the protection of the participating group. The machine-learning system identifies the group that tries to break into the system and disturbs the content of the system with this predefined pattern.

Machine learning and artificial intelligence must protect data quickly because hackers can access any device and obstruct intellectual property prior to an infringement occurring in the organisation. The assault is recognisable very early, and then neutralised, using artificial intelligence, to prevent the machine from further influencing it.

For an organisation focused on improving the cyber safety and mitigating the loss of confidential information, machine learning and artificial intelligence are an investment in several applications. Cybersecurity gets better every day with these resources.

Get start your journey with **Cyber Security Online Training**

Artificial intelligence offers powerful tools to improve cyber security algorithms performance. When tracing negative elements, the programme is extremely successful. Mechanisms for the monitoring of data are growing better and more reliable and reduce the relevant risks and improve operational performance. The cybersecurity room will benefit in many ways.

Cyber security professionals may assist in the study of high volume data sources and streams in several ways through machine learning and artificial intelligence.

Machine Learning Systems (ML) and Artificial Intelligence Systems (AI) for research

- Correlating different data sets in a particular pattern, scanning various risks, performing a predictive analysis and predicting the next attack.
- The constant monitoring of data security procedures may be carried out with the use of data purification techniques to protect users and other relevant parties and to ensure successful implementation of these restrictions.
- Build frameworks for data protection without the resource burden. Cybersecurity practitioners can minimise costs and prevent unnecessary expenses by machine learning and artificial intelligence.
- By building a security framework with integrated mechanisms to search huge volumes of data, data networks and identify any potential threats, various malware and infections can easily be identified with the help of artificial intelligence and machine learning.

Cloud storage is an essential factor in the description and determination of every company's online working patterns. Companies have migrated to different cloud systems, including Microsoft Azure that secure data more efficiently, from various hosting servers and equipment. Boon and bane come as each. This removes the use of hardware, which decreases the complexity of different operations. In the other side, a user's full faith is passed to a third party where no information in cloud providers is available first-hand, but all confidential information is now passed to others. This leaves the business vulnerable, but cloud storage is the best choice among its alternatives due to the absence of other choices. Software logarithms can easily be distributed across different cloud infrastructures with the aid of artificial intelligence.

2. IDENTIFYING ANOMALIES IN NETWORK TRAFFIC OR USER BEHAVIOR

Cisco expects a 6 percent annual compound growth rate (CAGR) from 3.9 billion Internet user worldwide in 2018 to 5.3 billion by 2023. But the most striking thing is that 10% CAGR is an increasing number of connexions and devices. By 2023, more than three times the world's population will be used to link computers to IP networks.

As the traffic increases, it becomes more difficult to detect anomalous activity in the network. AI and Machine Learn (ML) can quicker identifying behaviours that suggest potentials of danger and recognise deviations from regular patterns.

As with users, AI / ML may also examine the data stream for suspect behaviour suggesting an insider danger, such as uploading large quantities of data or trying to access unauthorised resources on a regular basis.

"The identification of networking defects will allow cyber security teams to locate from a corrupted hardware node to a disabled employee on the company network," writes John Burke from Nemertes Research at TechTarget.

Detecting online and other fraud

Just about 13% of organisations surveyed by the Certified Fraud Examiner Association have said they use computers to detect fraud, but they have tripled the amount of money they expect to investe by 2021. One case in point is that ML algorithms combine broad data groups such as historical data in order to find patterns far more efficiently than regulatory methods, which cybercriminals can subvert.

In order to decide what fraudulent activities and accounts look like compared to ordinary ones, for instance, supervised machine training will feed historical data. The model will then settle on account behaviour and characteristics. Instead, where there are few, or no "tagged" transactions, such as in modern forms of fraud, an unattended model may be used.

"You can identify previously unknown types of suspicious activity in an ideal combination of controlled and unmonitored AI techniques while identifying the most subtle patterns of fraud previously found across billions of accounts," writes TJ Horan on the FICO site.

Protecting endpoints

The second most popular answer to the issue, how AI enhances the protection status, was found in a ponemon / IBM safety survey by 2018. Most of the IT and safety professionals surveyed said that the greatest advantage of AI was rapid containment of contaminated devices and hosts.

Fast containment of a compromised host may mean an incomplete breakdown between failed cyber attacks. Machine learning techniques may help, for example, to detect legitimate malicious programme activity and stop a file from executing – effectively stop an attack in real time.

As noted in the latest annual IBM Security Data Infringement Report, 'the quicker and lower the cost of detecting and containing data infringement.' The report found that the average time for detecting and containing an incomplete infringement in 2019 increased to 279 days (from 266 in 2018). However, breaches with a life cycle with less than 200 days cost an estimated 1.22 million dollars less than average breaches of over 200 days (3.35 million dollars vs. 4.56 million dollars).

Mitigating the cybersecurity skills gap

The skill gap in cybersecurity continues to grow and impact on the ability of companies to fill vacancies — and thus effectively protect against attacks. Increasing the ability of a security team and automate threat detection and response will help, inter alia. AI / ML play a role in alleviating the shortage.

The Security Operations Center (SOC) is one area that could benefit. In its CISO Benchmark Report for 2020, Cisco concluded that 42 % of respondents encounter alert fatigue, primarily driven by a wide variety of security solutions.

A SOC analyst finds it time consuming to cut through noise on triage warnings as people are simply not cabled to effectively handle large numbers. This method can be greatly expanded by computer education and automation.

In an interview with VentureBeat BlackBerry CTO Charles Eagan put this way: "AI and automation are more about scalability than plugging unique capability gaps ... when we automatically eliminate 99% of the cyber threats, we can spend much more time and energy on ensuring security in more elaborate areas."

3. ARTIFICIAL INTELLIGENCE & MACHINE LEARNING IN CYBERSECURITY

Cyber security artificial intelligence – meanings of AI and ML

Artificial Intelligence (AI) and Machine Learning (ML) are extremely powerful, but they are only similar to the lack of clarification about their meaning.

We can find several concepts in AI, such as the powerful AI or true AI, which refer to artificial general intelligence, a hypothetical machine which shows at least as skillful and flexible behaviour as people.

In reality, a computer which can work and learn fully on its own outside of a regulated environment does not currently exist.

AI must be able to cope with large quantities of data, the ability to think, to organise and to arrange information that imitates human behaviour. It's mostly science fiction at this moment.

There is, however, general agreement that AI is an ML superset.

As a superset, AI has more subject than ML, but overlapping includes not just learning, such as speech recognition and comprehension, perception, imagination and intuition. Three dimensiones, 3D-understanding and interactions with the world. Commercial AI technologies involving AI additions over ML could be self-driving vehicles, computer vision and natural language processing (NLP).

Machine Learning is an AI discipline that enables computers to learn without being programmed directly. In essence, a programme for machine learning will find patterns in data and then preview the effects of something it has never before seen.

ML has been established by the latest advances in handling massive databases or big data, storage capability to support all these data and computer power.

ML are different types; supervised learning, profound learning and reinforcement learning are the most prevalent.

The majority of existing cyber-security AI applications do not surpass ML.

Artificial Intelligence in cybersecurity – Challenges to adopt ML

It is very important to be realistic about goals when introducing ML to incorporate any of the functions discussed previously. ML is always overviewed, and ML is driven by math and not by magic, we can not forget.

Data accessibility and consistency would probably be the toughest obstacle for adopting ML. We do not usually have all the information necessary in the proper context for feeding algorithms, such as appropriate attack data.

There is also a steep learning curve and major restrictions in the learning process.

The learning process should not start from scratch with enough context data, but it is not an easy task to provide this contextual data and exploit again.

When implementing an ML solution, we must ensure that we detect the right thing. The algorithms often do not learn the right thing. Furthermore, it is not easy to test and debug since there are many uncertainties we have to face.

In general, the highly specialised, limited and costly expertise requires important cost of acquisition, service and maintenance.

Regulation may be one last big obstacle or challenge. Regulatory mechanisms may have various impacts, including data security , data privacy and other policies that influence automated decision making.

Artificial Intelligence in cybersecurity – AI and ML used for evil

One thing is clear: AI and ML are resources and thus not bad or evil necessarily.

That said, since, as has previously been discussed, there are some fascinating applications to help the good guys, they might, and are, also being used to make havoc.

Definitely, the bad guys would try to use machine learning to help their attacks, learn from defensive responses and interrupt detection models.

In the use of machine learning, and advanated analytics, attackers need to expect more improvement than today with manual recognition techniques in accelerating and sharpering social engineering attacks such as phishing, fraud, DDos, ransomware, Spyware and scams across many business sectors.

In the case of ransomware, for example , attackers can use advanced analytics and ML to turn to more lucrative objectives, like high-net worth, IoT or special enterprises.

As we described earlier, cybersecurity machine speeds are crucial and hackers will make every effort to exploit newly discovered vulnerabilities faster than the defenders can fix.

Artificial Intelligence in cybersecurity – Ethics in ML

One of the first ethical problems surrounding ML undoubtedly concerns automation and the potential lack of human jobs. Given the current lack of expertise in the computer security industry, it is uncertain whether automation is as controversial as some claim.

When we consider predictive cyber protection for cyber crime or cyber terror – in which the defendants are interested in crimes that are not yet committed – there are additional ethical concerns. This approach comes directly into tension within the current legal system.

There are also possible problems caused by the low quality and/or insufficient amount of knowledge from which to base predictions and the predictive capacity to infer probabilistic outcomes of algorithms. Algorithmic transparency might pose a serious problem, particularly in controlled industries, and it is not easy to deal with this problem because it is not always feasible that everyone has access to the ML code will understand how the programme functions, mainly because ML algorithms don't work completely predictably.

Another very critical concern is that some of the acquired knowledge might be private or confidential. Under new regulations such as GDPR, this may be especially severe.

4. GENETIC ALGORITHM APPLIED TO INTRUSION DETECTION

Applying GA to intrusion detection seems to be a promising area. This paper discusses the motivation and implementation details in this section.

4.1 Overview

GAs can be used to evolve simple rules for network traffic. These rules are used to differentiate normal network connections from anomalous connections. These anomalous connections refer to events with probability of intrusions. The rules stored in the rule base are usually in the following form [4]:

if { condition } then { act } [4]

For the problems presented above, the condition usually refers to a match between current network connection and the rules in IDS, such as source and destination IP addresses and port numbers (used in TCP/IP network protocols), duration of the connection, protocol used, etc., indicating the probability of an intrusion. The act field usually refers to an action defined by the security policies within an organization, such as reporting an alert to the system administrator, stopping the connection, logging a message into system audit files, or all of the above. For example, a rule can be defined as:

if {the connection has following information: source IP address 124.12.5.18; destination IP address: 130.18.206.55; destination port number: 21; connection time: 10.1 seconds } then {stop the connection} [4]

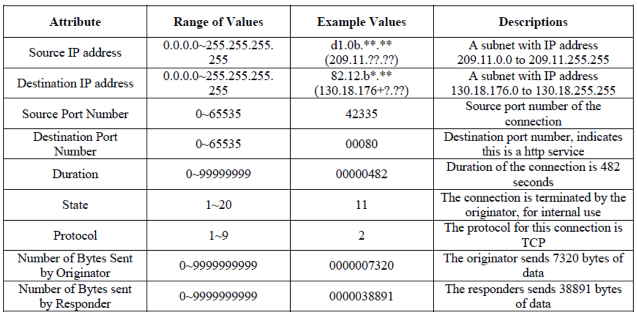
This rule can be explained as follows: if there exists a network connection request with the source IP address 124.12.5.18, destination IP address 130.18.206.55, destination port number 21, and connection time 10.1 seconds,

then stop this connection establishment. This is because the IP address 124.12.5.18 is recognized by the IDS as one of the blacklisted IP addresses; therefore, any service request initiated from it is rejected [4].

The final goal of applying GA is to generate rules that match only the anomalous connections. These rules are tested on historical connections and are used to filter new connections to find suspicious network traffic [4].

In this implementation, the network traffic used for GA is a pre-classified data set that differentiates normal network connections from anomalous ones. This data set is gathered using network sniffers (a program used to record network traffic without doing something harmful) such as Tcpdump or Snort. The data set is manually classified based on experts‘ knowledge. It is used for the fitness evaluation during the execution of GA. By starting GA with only a small set of randomly generated rules, we can generate a larger data set that contains rules for IDS. These rules are ―good enough‖ solutions for GA and can be used for filtering new network traffic.

Table 1. Rule definition for connection and range of values of each field



Attribute	Range of Values	Example Values	Descriptions
Source IP address	0.0.0.0~255.255.255.255	d1.0b.**.** (209.11.??.??)	A subnet with IP address 209.11.0.0 to 209.11.255.255
Destination IP address	0.0.0.0~255.255.255.255	82.12.b*.** (130.18.176+?.??)	A subnet with IP address 130.18.176.0 to 130.18.255.255
Source Port Number	0~65535	42335	Source port number of the connection
Destination Port Number	0~65535	00080	Destination port number, indicates this is a http service
Duration	0~99999999	00000482	Duration of the connection is 482 seconds
State	1~20	11	The connection is terminated by the originator, for internal use
Protocol	1~9	2	The protocol for this connection is TCP
Number of Bytes Sent by Originator	0~9999999999	0000007320	The originator sends 7320 bytes of data
Number of Bytes sent by Responder	0~9999999999	0000038891	The responders sends 38891 bytes of data

Altogether there are fifty-seven genes in each chromosome. For simplicity, this paper use hexadecimal representations for the IP addresses. The rule can be explained as follows: if a network connection with source IP address 209.11.??.?? (209.11.0.0 ~ 209.11.255.255), destination IP address 130.18.176.?? (130.18.176.0 ~ 130.18.255.255), source port number 42335, destination port number 80, duration time 482 seconds, ends with state 11 (the connection terminated by the originator), uses protocol type 2 (TCP), and the originator sends 7320 bytes of data, the responders sends 38891 bytes of data, then this is a suspicious behavior and can be identified as a potential intrusion [4]. The actual validity of this rule will be examined by matching the historical data set comprised of connections marked as either anomalous or normal. If the rule is able to find an anomalous behavior, a bonus will be given to the current chromosome. If the rule matches a normal connection, a penalty will be applied to the chromosome. Clearly no single rule can be used to separate all anomalous connections from normal connections. The population needs evolving to find the optimal rule set.

4.3 Parameters in Genetic Algorithm

There are many parameters to consider for the application of GA. Each of these parameters heavily influences the effectiveness of the GA. This paper will discuss the methodology and related parameters in the following section.

Evaluation function

The evaluation function is one of the most important parameters in GA. The proposed implementation differs from the scheme used by [3] in that the definition on calculations of **outcome** and **fitness** is different. The following steps are used to calculate the evaluation function.

First the overall **outcome** is calculated based on whether a field of the connection matches the pre-classified data set, and then multiply the weight of that field. The **Matched** value is set to either 1 or 0 [3].

$$Outcome = \sum_{i=1}^{57} Matched * Weight_i$$

$$\Delta = | outcome - suspicious\_level |$$

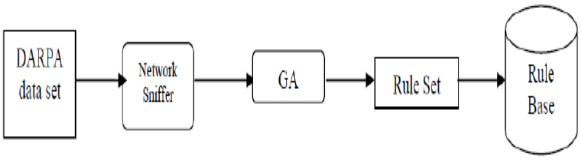


Figure 2: Architecture of applying GA into intrusion detection

Figure 2 shows the structure of this implementation. We need to collect enough historical data that includes both normal and anomalous network connections. The MIT Lincoln Laboratory data set for testing IDSs, which is represented in the Tcpdump binary format, is a good choice. This is the first part inside the system architecture. This data set is analyzed by the network sniffers and results are fed into GA for fitness evaluation. Then the GA is executed and the rule set is generated. These rules are stored in a database to be used by the IDS.

CONCLUSION

For all sorts of Internet users, be it people or large companies, cyber security has been the primary concern. The data is transmitted every day, every second, using different networks to expose it to various threats and risk data, which can not even be monetized. Cyber security becomes efficient and strong, but there is another side of the coin, with machine learning and artificial understanding coming into the picture. With the advent of machine learning and artificial intelligence, processes that expose systems to numerous threats become much simpler. Software breaking has become a play for children with AI and ML.

REFERENCES

[1] Neal Patwari, A. O. Hero III, M. Perkins, N. S. Correal, and R. O'Dea, "Relative location estimation in wireless sensor networks", IEEE Transactions and Signal Processing, Vol: 51, Issue: 8, 2003-8, Pages: 2137–2148.

[2] Jonathan Hui and David Culler, "The Dynamic Behavior of a Data Dissemination Protocol for Network Programming at Scale", In SenSys '04, ACM Conference on Embedded Networked Sensor Systems November 3-5, 2004. Baltimore, MD, In Proceedings of the 6th International Conference on Embedded Networked Sensor Systems.

[3] Erika Greene, Tugba Bodrumlu, and Kevin Knight, "Automatic analysis of rhythmic poetry with applications to generation and translation." In Proceedings of EMNLP, Conference on Empirical Methods in Natural Language Processing, October 9-11, 2010, MIT, Massachusetts, USA.

[4] Diana McCarthy, and Roberto Navigli, "Semeval-2007 task 10: English lexical substitution task." In Proceedings of the 4th International Workshop on Semantic Evaluations, Pages 48 to 53, 2007.

[5] Erika Odilia Flores Popoca, Maximino Miranda García, Socorro Romero Figueroa, Aurelio Mendoza Medellín, Horacio Sandoval Trujillo, Hilda Victoria Silva Rojas, Ninfa Ramírez Durán. "Pantoea agglomerans in Immunodeficient Patients with Different Respiratory Symptoms." The Scientific World Journal 2012, Article ID 156827, 8 pages.

[6] Christian Kehlmaier, Radek Michalko, Stanislav Korenko, "Ogcodes fumatus (Diptera: Acroceridae) Reared from Philodromus cespitum (Araneae: Philodromidae), and First Evidence of Wolbachia Alphaproteobacteria in Acroceridae, Annales Zoologici, 2012, 62:2, 281-286.

[7] M. Verbeek, A. M. Dullemans, P. J. van Bekkum, R. A. A. van der Vlugt, "Evidence for Lettuce big-vein associated virus as the causal agent of a syndrome of necrotic rings and spots in lettuce." Plant Pathology, Volume 62, Issue 2, pages 444–451, April 2013.

[8] Zhiwen Wang, Neil Hobson, Leonardo Galindo, Shilin Zhu, Daihu Shi, Joshua McDill, Linfeng Yang, Simon Hawkins, Godfrey Neutelings, Raju Datla, Georgina Lambert, David W. Galbraith, Christopher J. Grassa, Armando Geraldes, Quentin C. Cronk, Christopher Cullis, Prasanta K. Dash, Polumetla A. Kumar, Sylvie Cloutier, Andrew G. Sharpe, Gane K.-S. Wong, Jun Wang, Michael K. Deyholos, "The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads." The Plant Journal, 2012 Nov;72(3), P: 461-73.

[9] Raffaele Ronca, Michalis Kotsyfakis, Fabrizio Lombardo, Cinzia Rizzo, Chiara Currà, MartaPonzi, Gabriella Fiorentino, Josè M.C. Ribeiro, Bruno Arcà, "The Anopheles gambiae cE5, a tight- and fast-binding thrombin inhibitor with post-transcriptionally regulated salivary restricted expression", Insect Biochemistry and Molecular Biology, 2012, 42:9, 610-620.

[10] N.K. Suryadevara and S.C. Mukhopadhyay, “Wireless Sensor Network Based Home Monitoring System for Wellness Determination of Elderly”, IEEE Sensors Journal, Vol. 12, No. 6, June 2012, pp. 1965-1972.